

ENHANCEMENT OF GLOBAL TEMPO COMPUTATION IN BEAT TRACKING SYSTEM BASED ON TEAGER-KAISER ENERGY OPERATOR

Matěj Ištváněk

Master Degree Programme (2.), FEEC BUT

E-mail: xistva02@stud.feec.vutbr.cz

Supervised by: Zdeněk Smékal

E-mail: smekal@feec.vutbr.cz

Abstract: Beat detection systems and onset detections are used in music information retrieval (MIR) research field for the calculation of the global tempo (GT) and beat positions in audio recordings. The aim of this article is to introduce the enhancement of the onset detector and therefore the beat tracking system. The enhancement is based on the Teager-Kaiser energy operator (TKEO), which is used in pre-processing stage before the onset computation. The proposed method is firstly evaluated in terms of ability to estimate GT of a given audio track and then it is tested on the string quartet database. Results suggest that the TKEO could improve accuracy of GT estimation. Proposed beat tracking system could be used for analysis of interpretation changes in string quartet music.

Keywords: beat tracking, music analysis, music information retrieval, Teager-Kaiser energy operator, string quartet

1 ÚVOD

Vytvoření tempové a rytmické struktury hudby bez nutnosti poslechu samotných nahrávek je jedno z klíčových odvětví vědecké oblasti MIR (Music Information Retrieval). V ideálním případě systém dokáže přesně rozpoznat nejen celkové tempo skladby, ale také časový vývoj, časovou pozici beatů a kompletní rytmickou reprezentaci. Toho se využívá v tzv. autotagging systémech, doporučovacích algoritmech v rámci online hudebních streamovacích portálů nebo u klasifikačních algoritmů. Tempová struktura může vypovídat o povaze nahrávky a je důležitá také u hudebních analýz interpretačního výkonu. Spojení číslicového zpracování signálů a muzikologických výzkumů může vést k vytvoření nových tendencí na poli hudebního i technického vývoje.

2 NÁVRH A STRUKTURA ALGORITMU

Začátek nového tónu se v rámci hudebního signálu označuje jako „onset“. Pro určení tempové struktury je ve většině případů nutné nejprve vytvořit detektor tónů, resp. onsets. Existuje mnoho různých metod, které se liší především v samotném způsobu detekce nebo ve volbě časově-frekvenční reprezentace.

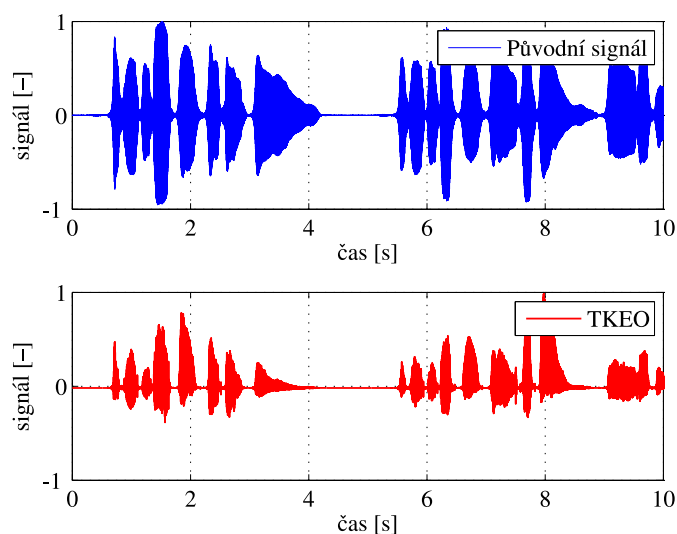
V navržené metodě byl vstupní signál v rámci předzpracování nejprve podvzorkován na 22050 Hz a převeden do mono formátu. Poté byl aplikován Teagerův-Kaiserův energetický operátor – TKEO ($\Psi[x[n]]$). Tento nelineární operátor má schopnost vytvořit energetickou obálku signálu, jak lze vidět na první testované nahrávce (10sekundovém úseku klarinetového sóla) – obr. 1. Na rozdíl od kvadratického operátoru může dosahovat záporných hodnot. Jeho definice je [1]:

$$\Psi(x[n]) = x^2[n] - x[n-1] \cdot x[n+1] \quad (1)$$

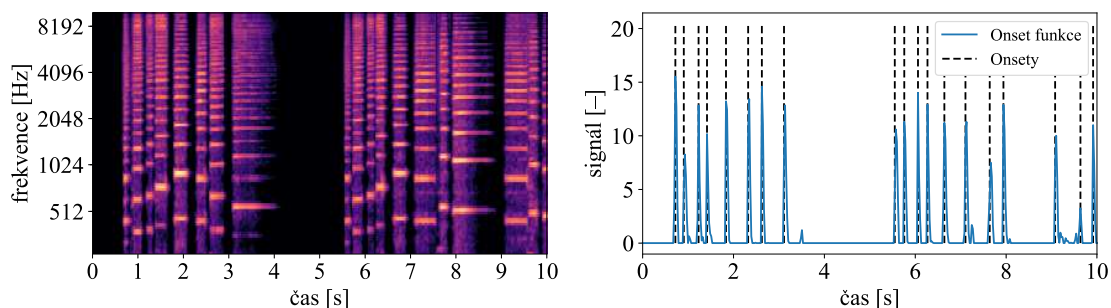
Následně byla vytvořena mel-frekvenční reprezentace, která využívá percepčního modelu lidského slyšení. Poté byl vypočítán spektrální tok (Spectral Flux) pro zjištění velikosti změn spektra mezi po sobě jdoucími segmenty [2]:

$$SF[m] = \frac{1}{K} \sum_{k=0}^{K-1} H(|X[m+1, k]| - |X[m, k]|) \quad (2)$$

pro $m = 0, 1, 2, \dots, M-2$, kde $H[x] = (x + |x|)/2$ je půlvlnný usměrňovač, M je počet segmentů a K je počet melovských pásem. Nakonec byla použita tzv. peak-picking funkce, která určí a následně označí časové pozice vzniklých onsetů pomocí adaptabilní prahové hodnoty a vytvoří výslednou detekční funkci (obr. 2). Pro tento účel byla použita knihovna programovacího jazyku Python – LibROSA¹.



Obrázek 1: Původní vstupní signál a TKEO první testované nahrávky.



Obrázek 2: Spektrogram a onset detekční funkce první testované nahrávky.

Nástroje, které vykazují nízkou změnu energie a spektra při změně tónu nebo hrají velmi rychle (například dvaatřicetinové tóny houslí v rychlém tempu), nebyly systémem správně detekovány a došlo ke snížení úspěšnosti algoritmu. Pro detekování globálního průměrného tempa celého segmentu (GT) u komplexnější polyfonické hudby nemusí být zapotřebí správná identifikace každého tónu. Důležitá je rytmická struktura a významné onsety.

¹<https://librosa.github.io/librosa/>

2.1 TEMPO

Pro vytvoření tempové struktury nahrávky potřebujeme zjistit, jak často se dané onsets opakují. Pro tento účel byl využit systém D. Ellise [3], který počítá autokorelační funkci vzniklé detekční funkce a hledá pravidelně se opakující onsets. TKEO byl použit v rámci předzpracování proto, aby pomohl detekčnímu systému vybrat vhodnější kandidáty pro správné pozice beatů.

Limitace studie spočívá v referenčním tempu nahrávek datasetu smyčcových kvartetů, které bylo odvozeno z notového zápisu a délky jednotlivých nahrávek – tedy nebere v potaz agogické a výrazové odchylky v rámci hudebního úseku, například hudební cézury nebo koruny. Navíc poměrně přesný výpočet GT ještě nemusí značit korektní časové pozice beatů v rámci rytmické výstavby skladby.

3 DATABANKA

Nejprve bylo testováno, zda systém s TKEO dosahuje lepších výsledků při detekci GT než stejný systém bez použití operátoru. Pro tento účel byla použita referenční SMC_MIREX databáze [4], která obsahuje 217 různých hudebních nahrávek s manuálně opravenými časovými pozicemi beatů.

Druhá testovaná databáze je tvořena 19 různými interpretacemi skladby Bedřicha Smetany *Smyčcový kvartet No. 1 e moll „Z mého života“*. První věta byla rozdělena do 6 logických segmentů (motivů).

4 VÝSLEDKY

Implementovaná metoda vykázala zlepšení v rámci porovnání přesnosti detekce GT nahrávek referenční databanky s konvenčním systémem. U prvních 50 nahrávek byla průměrná odchylka GT od referenční hodnoty pro systém bez použití TKEO 5,94 BPM (Beats Per Minute – beatů za minutu), což odpovídá 8,86 %. Pro navržený systém s TKEO 4,33 BPM (7,52 %). První motiv druhého datasetu vykázal podobné výsledky. Odchylka GT konvenčního systému od referenčního tempa byla v rozsahu 0,27–22,75 BPM (0,34–24,65 %) a u systému s použitím TKEO 0,27–14,36 BPM (0,34–15,56 %).

5 ZÁVĚR

U některých testovaných nahrávek v rámci detekce onsetů došlo ke zvýšení přesnosti systému, některé nahrávky vykázaly zhoršení detekce, a to především kvůli charakteru tvorby tónů jednotlivých nástrojů a stylu hudebního přednesu. Výsledky přesto naznačují, že použití TKEO v rámci předzpracování vstupního signálu může zvyšovat přesnost výpočtu GT. Další výzkum bude porovnávat přesnost detekce časových pozic beatů vůči jejich skutečným pozicím, aby bylo případně prokázáno nebo vyvráceno, že TKEO pomáhá algoritmu vybrat vhodnější onset kandidáty. Výsledky analýzy budou použity pro muzikologický výzkum rozdílů interpretace v závislosti na roku vzniku nahrávek.

REFERENCE

- [1] SMÉKAL, Z., T. KISKA a J. MEKYSKA. Teagerův-Kaiserův energetický operátor. *Sdělovací technika* č. 6, Praha, 2016, s. 10–13. ISSN 0036-9942.
- [2] MÜLLER, M. Fundamentals of Music Processing. NY: Springer Berlin Heidelberg, New York, 2015, s. 309–311. ISBN: 9783319219455. DOI: 10.1007
- [3] ELLIS, D. Beat Tracking by Dynamic Programming. *J. New Music Research, Special Issue on Beat and Tempo Extraction*, vol. 36 no. 1, 2007, s. 51–60. DOI: 10.1080/09298210701653344
- [4] HOLZAPFEL, A., M. E. P. DAVIES, J. R. ZAPATA, J. OLIVEIRA a F. GOUYON. Selective Sampling for Beat Tracking Evaluation. *IEEE Transactions on Audio, Speech and Language Processing*, vol. 20, no. 9, 2012, s. 2539–2548. DOI: 10.1109/TASL.2012.2205244